

APPENDIX A. TECHNICAL EXPERT PANEL

Edward A. Chow, MD (retired)

Medical Director, Chinese Community Health Plan
Executive Director, Chinese Community Health Care Association
San Francisco Health Commission

Cheryl Damberg, PhD, MPH

Senior Principal Researcher, Professor
Pardee RAND Graduate School

Laura Damschroder, MS, MPH

Research Scientist, Ann Arbor VA Center for Clinical Management Research
Co-implementation Research Coordinator, Diabetes QUERI
Ann Arbor VA HSR&D Center of Excellence

Laura Dimmler, MPA, PhD

Director, School of Healthcare Administration and Leadership
Associate Professor, Pacific University

John McConnell, PhD

Director, Center for Health Systems Effectiveness & Associate Professor, Department of
Emergency Medicine, Department of Public Health & Preventative Medicine, and
Division of Management, OHSU

Richard Stenson, MHA, MBA (retired)

President & CEO, Tuality HealthCare and Tuality Health Alliance

Kevin Volpp, MD, PhD

Director, Center for Health Incentives and Behavioral Economics, Leonard Davis Institute
Director, Penn CMU Roybal P30 Center in Behavioral Economics and Health

APPENDIX B. PICOTS TABLE

	<p>KQ1. Effectiveness of pay for performance on patient outcomes and processes of care.</p> <p>a. What are the effects of pay for performance programs on patient outcomes and process of care?</p> <p>b. Are there certain intervention characteristics (<i>ie</i>, size of incentive, target of incentive) that are associated with beneficial effects of these programs?</p> <p>c. For which populations of patients are financial incentive programs most effective?</p>	<p>KQ2. Implementation factors.</p> <p>Which implementation factors modify the effectiveness of pay for performance?</p>	<p>KQ3. Unintended consequences and harms.</p> <p>a. What are the positive unintended consequences related to pay for performance?</p> <p>b. What are the negative unintended consequences related to pay for performance?</p> <p>c. What is the effect of pay for performance on inequality/health disparities?</p> <p>d. What are the intervention and implementation factors that contribute to or mitigate the positive and negative unintended consequences of financial incentive programs?</p>
Target population	Healthcare providers at the individual, managerial (eg, VISN directors), group, and institutional levels. General patient populations that are part of existing performance measures.		
Intervention	Financial incentives/pay-for performance programs		
Comparator	Other financial incentive models; non-financial incentives; usual care Examples of factors to examine or compare: <ul style="list-style-type: none"> - individual vs provider groups vs institutions - patient outcomes vs processes of care - structure of the incentive (eg, relationally determined or can everyone receive award?) - size of the incentive - target patient population (chronic illness vs disease specific) - how the payment is made (bonus vs salary) - duration of the incentive - positive vs negative incentives - other implementation factors 		
Outcomes	<p>A. Performance measures in patients</p> <ul style="list-style-type: none"> - quality-of-life measures - mortality and morbidity - health care utilization (eg, admissions, ER visits) - intermediate physiological markers such as blood pressure, HbA1c, and cholesterol - health promotion outcomes such as smoking cessation, alcohol/substance abuse, and weight loss <p>B. Processes of care</p> <ul style="list-style-type: none"> - Access to care - Preventive screening - Referral - Health behavior education 		Unintended consequences and associated cognitive processes such as motivation (extrinsic vs intrinsic motivation), gaming, risk selection, spillover effects. In addition, unintended consequences may relate to the exacerbation of health disparities in low-income and ethnic minority populations.
Timing	?		
Setting	VHA or other large managed care institutions, other healthcare systems in the US, and healthcare systems in countries with health systems similar to the VHA.		
Study designs	Studies with concurrent controls.		All study designs will be considered

APPENDIX C. SEARCH STRATEGIES

PubMed Searched April 3, 2014

Search String	Notes
("Reimbursement, Incentive"[Mesh]) OR "Physician Incentive Plans"[Mesh]	Mesh Terms for a specific search of indexed articles
(((publisher[sb]) OR inprocess[sb]) OR pubmednotmedline[sb]) OR oldmedline[sb] AND (((((((((((((((((((((((("pay for performance"[Title/Abstract]) OR p4p[Title/Abstract]) OR pfp[Title/Abstract]) OR "pay for value"[Title/Abstract]) OR "payment for quality"[Title/Abstract]) OR "performance-based payment"[Title/Abstract]) OR "performance-based reimbursement"[Title/Abstract]) OR "performance-based contracting"[Title/Abstract]) OR "performance-based pay"[Title/Abstract]) OR "output-based payment"[Title/Abstract]) OR "incentive reimbursement"[Title/Abstract]) OR "incentive program"[Title/Abstract]) OR "quality based purchasing"[Title/Abstract]) OR "quality incentive"[Title/Abstract]) OR "quality incentives"[Title/Abstract]) OR "quality payment"[Title/Abstract]) OR "quality payments"[Title/Abstract]) OR "quality-based payment"[Title/Abstract])) OR (("financial incentive"[Title/Abstract]) AND effectiveness[Title/Abstract])) OR (("financial incentives"[Title/Abstract]) AND effectiveness[Title/Abstract])) OR (("monetary incentive"[Title/Abstract]) AND effectiveness[Title/Abstract])) OR (("monetary incentives"[Title/Abstract]) AND effectiveness[Title/Abstract])) OR ((bonus[Title/Abstract]) AND "quality"[Title/Abstract]) OR (("reward"[Title/Abstract]) AND "quality"[Title/Abstract]) OR (("quality based"[Title/Abstract]) AND payments[Title/Abstract]))	Keyword terms for a sensitive search of non-indexed articles
Keyword search and MeSH search combined with OR	
Limited to publication date after 07/07/2011	Date of Eij. Search

Pay-For-Performance Literature Review

Search String	Notes
PubMed Searched from December 2012 to current Searched on April 30, 2014	
(((("pay for performance"[Title/Abstract]) OR P4P[Title/Abstract]) OR "pay for value"[Title/Abstract]) OR "financial incentive"[Title/Abstract])) OR (((bonus OR reward[Title/Abstract])) AND (payment OR reimburse* OR incentive*[Title/Abstract])) AND (quality OR value[Title/Abstract]))	[exact copy of Rand Search] Saved as "P4P Rand Gap Search 1"
((((((((((Beckman, Howard[Author]) OR Curtin, Kathleen[Author]) OR Casalino, Larry[Author])	[author search copy of Rand] Saved as "P4P Rand Gap Search 2"

OR Dudley, Adams[Author]) OR Doran, Tim[Author]) OR Jha, Ashish[Author]) OR Petersen, Laura[Author]) OR Roland, Martin[Author]) OR Rosenthal, Meredith[Author]) OR Ryan, Andrew[Author]) OR Schneider, Eric[Author]) OR Werner, Rachel[Author]) OR Damberg, Cheryl[Author]	authors” After deduplication with search 1, 204 unique results
Searches in Additional Databases are from June 2007 to current Searched on April 30, 2014	
CINAHL (EBSCO)	
Exact copy of above search strategy (no subject heading used, so no translation needed)	N=1559 After deduplication with PubMed Searches 1319 unique results
PsycInfo(Ovid)	
Exact copy of above search strategy (no subject heading used, so no translation needed)	N=1183 After deduplication with PubMed and Cinahl searches 1177 unique results
	The Pay for Performance Literature Review by Rand also searched EconLit and ABIInform we do not have access to either of these databases.

Accountable Care Organization Literature Review

Search String	Notes
Pubmed Searched from November 2012 to current Searched on April 30 th 2014	
(((((quality[Title/Abstract]) OR quality improvement) OR quality indicators, health care) OR "quality of care") OR "quality of healthcare")) AND (((accountable care organization*) OR ACO) OR ACOS)	[exact copy of Rand search] N=129 Saved as “ACO Rand Gap Search”
Medline (OVID) [Ovid MEDLINE® without Revisions 1996-April Week 3 2014 ; Ovid MEDLINE ® and Ovid OLDMEDLINE ® 1946 to April Week 3 2014 ; Ovid MEDLINE ® In-Process & Other Non-Indexed Citations April 29,2014] Searched April 30, 2014 and limited to 2012 to current	
(share\$ adj3 savings).mp.	[exact copy of Rand search] N=53 After deduplication with search above 37 unique results
((accountable adj2 care adj2 organization\$.mp OR (ACO OR ACOS).mp. NOT (gene OR genetics\$.mp.) AND (Algorithms\$.mp OR algorithms/)	N=20 after deduplication with above searches 1 (one) unique citation

Search of WorldCat limited to November 2012 to current searched on April 30, 2014

Search String	Notes
'((kw: accountable and kw: care and kw: organization* OR kw: aco OR kw: acos)) OR ((kw: shared and kw:saving*)) and 9kw: health* OR kw: medical OR kw: patient* OR kw: physician* OR kw: doctor* OR kw: hospital* OR kw: nurs*)'	[exact copy of rand search]

APPENDIX D. INCLUSION AND EXCLUSION CRITERIA

1. Language: Is the full text of the article in English?	
Yes.....	Proceed to #2
No	Code X1 . STOP
2. Population: Is the population human participants?	
Yes	Proceed to #3
No	Code X2 . Add code B if retaining for background/discussion. STOP
3. Financial Incentives Intervention: Does the article include information relevant to financial incentive programs?	
Yes	Proceed to #4
No	Code X3 . Add code B if retaining for background/discussion. STOP
4. Financial Incentives Setting: Does the article assess pay for performance programs or accountable care organizations in a healthcare setting? Other settings such as businesses or education are excluded. <i>Note: Common incentive programs are the Quality and Outcomes Framework (QOF) and the Hospital Quality Incentive Demonstration (HQID), Advancing Quality (AQ), Clalit P4P, Clinical Practice Improvement Payment (CPIP), Ergebnis Orientierte Vergütung (EOV), Maccabi P4P, National Health Insurance P4P, Performance Management Program (PMP), Physician Integrated Network (PIN), Practice Incentive Program (PIP), Primary Care P4P, Primary Care Renewal Models (PCRM), Program of Quality Improvement (PQI), the Premier Demonstration, the Physician Group Practice Demonstration, the Integrated Healthcare Association P4P program, the Blue Cross Hawaii P4P program, the Massachusetts multi-plan P4P program, and the Blue Cross Blue Shield of Massachusetts AQC. Common ACOs are the CMS ACO demonstration, the Medicare Pioneer ACO, and the Pioneer ACO.</i>	
Yes	Proceed to #5
No	Code X4 . Add code B if retaining for background/discussion. STOP
5. Financial Incentives Population: Does the article assess direct financial incentives or pay for performance programs targeting healthcare providers at the individual, managerial, group, institutional, or system level? Financial incentives targeting patient populations are excluded.	
Yes	Proceed to #6
No	Code X5 . Add code B if retaining for background/discussion. STOP
6. Financial Incentives Population: Does the article assess direct financial incentives or pay for performance programs targeting healthcare providers at system level (e.g., capitation, managed care, bundled payments)?	
Yes.....	Code X6 . Add code B if retaining for background/discussion. STOP
No.....	Proceed to #7
7. Study Design: Is the study design a randomized controlled trial?	
Yes.....	Code T . STOP
No.....	Proceed to #8
8. Study Design: Is the study design a review (systematic, literature, meta-analysis)?	
Yes.....	Code R . STOP
No.....	Proceed to #9
9. Study Design: Is the study design observational?	
Yes.....	Code O . STOP
No.....	Proceed to #10
10. Study Design: Is the study design a case study, case series, or case report?	

<p>Yes..... Code C. STOP</p> <p>No.....Proceed to #11</p> <p>11. Study Design: Is the study design qualitative?</p> <p>Yes..... Code Q. STOP</p> <p>No.....Proceed to #12</p> <p>12. Study Design: Is the article a commentary, letter to the editor or editorial?</p> <p>Yes..... Code E. STOP</p> <p>No.....Proceed to #13</p> <p>13. Study Design: All other study designs, or if the study design is unclear..... Code U. STOP</p>
<p>Key Question 1:</p> <ul style="list-style-type: none"> d. What are the effects of financial incentive programs on patient outcomes and process of care? e. Are there certain intervention characteristics (<i>ie</i>, size of incentive, target of incentive) that are associated with beneficial effects of these programs? f. For which populations of patients are financial incentive programs most effective? <p>Key Question 2: Which implementation factors modify the effectiveness of financial incentives?</p> <p>Key Question 3:</p> <ul style="list-style-type: none"> e. What are the positive unintended consequences related to financial incentives? f. What are the negative unintended consequences related to financial incentives? g. What is the effect of financial incentives on inequality/health disparities? h. What are the intervention and implementation factors that contribute to or mitigate the positive and negative unintended consequences of financial incentive programs?

APPENDIX E. STUDIES SUMMARIZED IN DAMBERG, 2014¹

1. Campbell S, Reeves D, Kontopantelis E, Middleton E, Sibbald B, Roland M. Quality of primary care in England with the introduction of pay for performance. *New England Journal of Medicine*. 2007;357(2):181-190.
2. Chien AT, Li Z, Rosenthal MB. Improving timely childhood immunizations through pay for performance in Medicaid-managed care. *Health Services Research*. 2010 Dec;45(6 Pt 2):1934–1947.
3. Chien AT, Eastman D, Li Z, Rosenthal MB. Impact of a pay for performance program to improve diabetes care in the safety net. *Preventive Medicine*. 2012 Nov;55 Suppl:S80–S85.
4. Christianson J, Leatherman S, Sutherland K. Lessons from evaluations of purchaser pay-for-performance programs: A review of the evidence. *Medical Care Research and Review*. 2008;65(6 Suppl):5S–35.
5. Fagan PJ, Schuster AB, Boyd C, Marsteller JA, Griswold M, Murphy SM, Dunbar L, Forrest CB. Chronic care improvement in primary care: Evaluation of an integrated payfor-performance and practice-based care coordination program among elderly patients with diabetes. *Health Services Research*. 2010 Dec;45(6 Pt 1):1763–1782..
6. Glickman SW, Ou FS, DeLong ER, Roe MT, Lytle BL, Mulgund J, Rumsfeld JS, Gibler WB, Ohman EM, Schulman KA, Peterson ED. Pay for performance, quality of care, and outcomes in acute myocardial infarction. *JAMA*. 2007 Jun 6;297(21):2373–2380.
7. Krumholz H, Lin Z, Keenan P, Chen J, Ross J, Drye E, Bernheim S, Wang Y, Bradley E, Han L, Normand S. Relationship between hospital readmission and mortality rates for patients hospitalized with acute myocardial infarction, heart failure, or pneumonia. *JAMA*. 2013;309(6):587–593.
8. Levin-Scherz J, DeVita N, Timbie J. Impact of pay-for-performance contracts and network registry on diabetes and asthma HEDIS measures in an integrated delivery network. *Medical Care Research and Review*. 2006 Feb;63(1 Suppl):14S–28S.
9. Lindenauer PK, Remus D, Roman S, Rothberg MB, Benjamin EM, Ma A, Bratzler DW. Public reporting and pay for performance in hospital quality improvement. *New England Journal of Medicine*. 2007;356(5):486–496.
10. Mullen KJ, Frank RG, Rosenthal MB. Can you get what you pay for? Pay-for performance and the quality of healthcare providers. *Rand Journal of Economics*. 2010 Spring;41(1):64–91.
11. Nicholas LH, Osborne NH, Birkmeyer JD, Dimick JB. Hospital process compliance and surgical outcomes in medicare beneficiaries. *Archives of Surgery*. 2010 Oct;145(10):999–1004.
12. Pearson SD, Schneider EC, Kleinman KP, Coltin KL, Singer JA. The impact of pay-for-performance on health care quality in Massachusetts, 2001–2003. *Health Affairs*. 2008 Jul–Aug;27(4):1167–1176.
13. Petersen LA, Woodard LD, Urech T, Daw C, Sookanan S. Does pay-for-performance improve the quality of health care? *Annals of Internal Medicine*. 2006 Aug 15;145(4):265–272.

14. Rosenthal MB, Li Z, Robertson AD, Milstein A. Impact of financial incentives for prenatal care on birth outcomes and spending. *Health Services Research*. 2009 Oct;44(5 Pt 1):1465–1479.
15. Ryan A. Effects of the Premier Hospital Quality Incentive Demonstration on Medicare patient mortality and cost. *Health services research*. 2009;44(3):821–842.
16. Ryan AM, Blustein J. The effect of the MassHealth hospital pay-for-performance program on quality. *Health services research*. 2011;46(3):712-728.
17. Ryan AM, Doran T. The effect of improving processes of care on patient outcomes: Evidence from the United Kingdom's quality and outcomes framework. *Medical care*. 2012 Mar;50(3):191–199.
18. Ryan AM, Blustein J, Doran T, Michelow MD, Casalino LP. The effect of Phase 2 of the Premier Hospital Quality Incentive Demonstration on incentive payments to hospitals caring for disadvantaged patients. *Health Services Research*. 2012 Aug;47(4):1418–1436.
19. Sidorenkov G, Haaijer-Ruskamp FM, de Zeeuw D, Bilo H, Denig P. Review: Relation between quality-of-care indicators for diabetes and patient outcomes: A systematic literature review. *Medical Care Research and Review*. 2011 Jun;68(3):263–289.
20. Sutton M, Nikolova S, Boaden R, Lester H, McDonald R, Roland M. Reduced mortality with hospital pay for performance in England. *New England Journal of Medicine*. 2012 Nov 8;367(19):1821–1828.
21. Van Herck P, De Smedt D, Annemans L, Remmen R, Rosenthal MB, Sermeus W. Systematic review: Effects, design choices, and context of pay-for-performance in health care. *BMC health services research*. 2010;10:247-247.
22. Werner RM, Bradlow ET. Relationship between Medicare's hospital compare performance measures and mortality rates. *JAMA*. 2006 Dec 13;296(22):2694–2702.
23. Werner RM, Kolstad JT, Stuart EA, Polsky D. The effect of pay-for-performance in hospitals: Lessons for quality improvement. *Health Affairs*. 2011 Apr;30(4):690–698.

APPENDIX F. KEY INFORMANT DISCUSSION GUIDE, TEMPLATE

Portland Evidence-based Synthesis Program

Understanding the intervention and implementation factors associated with benefits and harms of pay for performance programs in healthcare

Dr. <KEY INFORMANT>

<MONTH, DAY, 2014: TIME PT/ ET>

Conference call: 1.800.767.1750, Access Code: 39528#

1. According to your study, financial incentives had XXX effect on XXXX. What were some of the main factors that contributed to your results?
 - a. Probe: Intervention variables such as:
 - i. Rewards vs penalties
 - ii. Type/nature
 - iii. Relative vs absolute performance measures
 - iv. Frequency and duration
 - v. Certainty of the incentive
 - b. Probe: Implementation factors such as:
 - i. Inner setting (structural, political, cultural contexts)
 - ii. Outer setting (economic, political, social contexts)
 - iii. Individuals involved (cultural, organizational, professional, and individual mindsets, norms, interests, affiliations)
 - iv. Implementation processes (interaction of related processes within the organization)
2. What did you find were some of the unintended consequences related to financial incentives?
 - a. Probe: Positive
 - i. Spillover effects
 - b. Probe: Negative
 - i. Risk selection
 - ii. Deterioration of un-incentivized care
 - iii. Impairment of intrinsic motivation/professionalism
 - iv. Gaming
3. Did you find that financial incentives had any effect on health disparities?
 - a. Probe: Were there certain groups that were at a greater disadvantage?
 - i. Low income
 - ii. Racial/ethnic minority populations
 - b. Probe: Why?
 - i. Access
 - ii. Language barriers
 - iii. Lack of insurance/ability to pay
 - iv. Etc.
4. What were some of the things that were the most surprising to you?
5. What would you have done differently?

APPENDIX G. KEY INFORMANTS

Howard Beckman, MD, FACP, FAACH
CMO, Focused Medical Analytics
Clinical Professor of Medicine, Family
Medicine and Public Health Sciences
University of Rochester School of Medicine
and Dentistry

Justin Benzer, PhD
Research Health Scientist & Research
Assistant Professor, Center for Healthcare
Organization and Implementation Research
VA Boston Healthcare System & Boston
University

Sule Calikoglu, PhD
Maryland Health Services Cost Review
Commission

Alyna T. Chien, MD, MS
Harvard Medical School
Boston Children's Hospital

**Tim Doran, BSc, MBChB, MPH, MD,
MFPH**
Professor of Health Policy, University of York

Peter J. Fagan, PhD, MDiv.
Associate Professor of Medical Psychology
Johns Hopkins University School of Medicine

**Rachel Foskett-Tharby, PhD, MSc, BSc,
RGN**
Research Fellow
University of Birmingham

Eve A. Kerr, MD, MPH
Director, VA Center for Clinical Management
Research

Lauren Hersch Nicholas, PhD, MPP
Johns Hopkins Bloomberg School of Public
Health & School of Medicine, Department of
Health Policy & Management and Department
of Surgery

Armando Henrique Norman, MD
Department of Anthropology
Durham University

Laura A. Petersen, MD, MPH, FACP
MEDVAMC Associate Chief of Staff,
Research
Director, VA HSR&D Center for Innovations
in Quality, Effectiveness & Safety (IQeSt)

Martin Roland, CBE, DM, FMedSci
RAND Professor of Health Services Research,
Institute of Public Health
University of Cambridge School of Clinical
Medicine

Andrew M Ryan, PhD
Division of Outcomes and Effectiveness
Research
Weill Cornell Medical College

Rachel Werner, MD, PhD
Center for Health Equity Research and
Promotion
Philadelphia VAMC
Associate professor of Medicine
University of Pennsylvania

APPENDIX H. PEER REVIEW COMMENTS AND RESPONSES

Question Text	Comment	Response
Are the objectives, scope, and methods for this review clearly described?	Yes	
	Yes	
Is there any indication of bias in our synthesis of the evidence?	No	
	No	
Are there any published	No	

<p>or unpublished studies that we may have overlooked?</p>	<p>Yes - For consideration. Not all of these may be directly/exclusively related to P4P but may provide context.</p> <ul style="list-style-type: none"> · Medicare's public reporting initiative on hospital quality had modest or no impact on mortality from 3 key conditions AM Ryan, BK Nallamothu, JB Dimick - Health Affairs, 2012 – · Has Pay-for-Performance Decreased Access for Minority Patients? AM Ryan - Health services research, 2010 · The long-term effect of premier pay for performance on patient outcomes AK Jha, KE Joynt, EJ Orav, AM Epstein - New England Journal of Medicine, 2012 · Medicare's flagship test of pay-for-performance did not spur more rapid quality improvement among low-performing hospitals AM Ryan, J Blustein, LP Casalino - Health affairs, 2012 · The unintended consequences of publicly reporting quality information RM Werner, DA Asch - Jama, 2005 · Does hospital performance on process measures directly measure high quality care or is it a marker of unmeasured care? RM Werner, ET Bradlow, DA Asch - Health Services Research, 2008 · Making the 'pay' matter in pay-for-performance: Implications for payment strategies RM Werner, RA Dudley - Health Affairs, 2009 · Effects of pay for performance in health care: A systematic review of systematic reviews F Eijkenaar, M Emmert, M Scheppach, O Schöffski - Health Policy, 2013 · Early experience with pay-for-performance: from concept to practice MB Rosenthal, RG Frank, Z Li, AM Epstein 	<p>Thank you for the list of additional articles. As mentioned, many of those listed are either included in the RAND report, thus not included in our report, or are not directly related to P4P; however, do provide context/background.</p> <ul style="list-style-type: none"> · Ryan, Nallamothu, et al (2012) examines the effect of public reporting on mortality. · Ryan (2010) is included in the RAND report. · Jha et al (2012) is included in the RAND report. · Ryan, Blustein et al (2012) is included in the RAND report. · Werner & Asch (2005) is a great background piece on public reporting. · Werner et al (2008) provides good background on the relationship between process measures and outcomes. · Werner & Dudley (2009) is a study examining different payment strategies. We have included this paper in the revision (KQ2). · Eijkenaar et al (2013) is a systematic review of reviews, and does not meet inclusion criteria based on study design; however, we did reference this paper in our background. · Rosenthal et al (2005) is included in the RAND report.
--	--	---

	<p>Yes - I did not find CMS, Meaningful Use for EHR, or NCQA data which would reflect the P4P programs for Medicare and for health plans. These are also P4P programs. Another source for value of quality programs is AHIP. The CMS programs now have reduction of payments for hospitals, and now also have take back for deficiencies including the DSRIP program. The Meaningful use program has impacted many individual physicians. NCQA has literature concerning improved quality -- and quality programs of plans for providers.</p>	<p>Thank you very much. The literature included in our report included only studies that were published and/or not included in the RAND report, and it is possible that some of the research related to the mentioned programs were excluded due to search date limitations. In addition we limited our scope to programs that were direct P4P programs, and did not include those that were ACOs or bundled payments. In response to your review comment, our research librarian conducted a search of the mentioned organization/program websites for unpublished studies meeting our inclusion criteria. None were located.</p>
	<p>Yes - consider adding Rachel Werner on Denominator Gaming in NH Compare (not a P4P study but does have financial implications</p>	<p>Thank you. This study does not speak specifically to P4P; thus, did not meet inclusion criteria. However, we have added a statement in future research needs calling for explicit research examining negative unintended consequences related to P4P, including denominator gaming.</p>
	<p>No</p>	
	<p>No</p>	

	<p>Yes –</p> <ul style="list-style-type: none"> • Blustein, J et al (2011). Analysis raises questions on whether pay-for-performance in Medicaid can efficiently reduce racial and ethnic disparities. <i>Health Affairs</i>, 30 (6), pp. 1165-1175. • McHugh, M.D., et al. (2010). Medicare readmissions policies and racial and ethnic health disparities: A cautionary tale. <i>Policy, Politics, & Nursing Practice</i>, 11 (4), pp. 309-316. • Lewis, V.A., et al. (2012). The promise and peril of accountable care for vulnerable populations: A framework for overcoming obstacles. <i>Health Affairs</i>, 31 (8), pp. 1777-1785. • Hearld, L.R. et al. (2014). Pay-for-performance and public reporting program participation and administrative challenges among small- and medium-sized physician practices. <i>Medical Care Research and Review</i>, 71(3), pp. 299-312. • Casalino, L.P., et al. (2007). Will pay-for-performance and quality reporting affect health care disparities? <i>Health Affairs</i>, 26 (3), pp. 405-414. • Chien, A.T., et al. (2007). Pay-for-performance, public reporting, and racial disparities in health care: How are programs being designed? <i>Medical Care Research and Review</i>, 64 (5), pp. 283-304. • Crawley, D., et al. (2009). Impact of pay-for-performance on quality of chronic disease management by social class group in England. <i>Journal of the Royal Society of Medicine</i>, 102, pp. 103-107. • Weinick, R.M., et al. (2011). Quality improvement efforts under health reform: How to ensure that they help reduce disparities. <i>Health Affairs</i>, 30 (10), pp. 1837-1843. 	<p>Thank you for providing a list of articles.</p> <ul style="list-style-type: none"> • Blustein et al (2011). Although the article discusses the program in light of P4P, the data that they present did not represent measures that were incentivized at the time . Our inclusion criteria limits inclusion to studies that include incentivized measures. • McHugh et al (2010) provides great background on racial disparities in hospital readmissions; however, it does not evaluate outcomes related to P4P. • Lewis et al (2012) provides a framework for considering vulnerable populations in ACOs. Our scope was limited to studies of primary data examining P4P, as distinguished from ACOs. • Hearld et al (2014) is included in KQ2. • Casalino et al (2007) is a good background paper with recommendations; however, provides no data for inclusion in the systematic review. • Chien et al (2007) is a systematic review and qualitative study examining health disparities. Systematic reviews and qualitative studies were not included in our review. The recommendations identified from their program leader interviews are congruent with our KI interviews. We will reference this in our revision. • Crawley et al (2009) is included in KQ3. • Weinick et al (2011) provides recommendations for reducing health disparities; however, does not evaluate outcomes related to P4P.
--	--	--

<p>Additional suggestions or comments can be provided below. If applicable, please indicate the page and line numbers from the draft report.</p>	<p>See attached comments: General Comments: 1. This report bases findings on a combination of a recent high-quality systematic review and an updated literature search that located 93 additional studies. The approach is justifiable and efficient, and the findings are presented in a reasonably clear text that distinguishes conclusions from the RAND report from the additional studies. The report seems sounds in its conclusions and appropriate for policy makers. However, it is not easy to compare the relative weight of evidence from the RAND report vs the additional studies. Some mention of the number of studies included in the RAND report or the total number of patients might help – <i>ie</i>, does the updated evidence since 2012 more than double the amount of evidence reviewed in the RAND report? 2. The inclusion of key informant interviews is a welcome addition, although the methods are described in a very limited fashion. A little more detail on how themes were extracted would be helpful (eg, were interviews recorded and transcribed, was any qualitative software used, did multiple people analyze same interview, etc). 3. The executive summary should get a careful proofreading. I noted several minor grammatical errors (a missing or extra word on p. 1 line 27; p. 3 line 25) 4. The concept of “increasing maximum thresholds” wasn’t clear to me – does this mean setting a higher target for P4P – eg, 90% vs 80% attainment ? Please state more clearly as it might be construed as maximum payment. 5. Please explain concept of “penalties” vs “rewards” – I assume you mean the idea of withholds on reimbursement (or placing a % of capitation at risk). 6. On the answers under key Question 2, is it possible to include any more specific qualifiers than “studies” – this could mean 2 studies or 6 studies, and it isn’t clear if any studies found opposing results. A clearer introduction might say – “among findings that were consistently reported by more than one study...” if that is what the observations represent. If some findings appear more robust, --<i>ie</i>, the result reported by the most studies – that should be noted and reported first. Otherwise it is hard to distinguish what might be relatively anecdotal evidence vs more compelling findings. 7. Some of the policy implications could be more specific if there is evidence to be gleaned from the studies – for example, what designs would mitigate gaming? What is a reasonable # of</p>	<ol style="list-style-type: none"> 1. Thank you. Distinguishing the relative weight of the evidence presented in the RAND report vs this report is challenging, as our inclusion criteria were different. Because RAND’s report was commissioned by CMS, with the exception of a few studies examining health disparities and one study looking at the link between process and intermediate measures, they did not include studies of the Quality and Outcomes Framework (QOF) due to differences in the health systems. Conversely, the QOF was suggested by one of our stakeholders as being the P4P program in a system that was most similar to the VHA; thus, not only did we include studies related to the QOF in our search strategy, we also conducted targeted searches for both published and unpublished literature for findings related to the QOF. As our primary literature search began at the end date of RAND’s search and since we excluded all studies published in RAND’s report, new evidence associated with programs other than the QOF was limited. However based on our primary and targeted searches, we included a total of 47 studies examining the QOF. Another significant difference between RAND’s report and this report is that due to our large number of included QOF studies, we present mostly findings associated with P4P in ambulatory settings (78 studies), with only 11 studies examining P4P in hospital settings. RAND’s report included 48 studies conducted in ambulatory settings, and 38 examining P4P in hospital settings. We have revised the report to include the total number of studies included in the RAND report along with a breakdown of number of studies associated with ambulatory and hospital settings. Our revision also includes the total number of studies in each ambulatory and hospital settings that were included in this report. We have also added to the discussion/limitations a mention of the large number of QOF and ambulatory studies, and limited number of studies conducted in hospital settings. 2. Thank you. We have updated the methods section in both the executive summary and the main report to include a more detailed description. 3. Thank you 4. Thank you. We added a definition to the first instance of the term in both the executive summary and the main report. 5. Thank you. There are variety of ways in which penalties may be applied, such meeting targets to earn withholds, as well as repayments to payers for failure to meet benchmarks. We
--	---	--



	<p>measures to be sufficiently broad but not overburdening. 8. Given number of studies on QOF, a longer introductory description of the nature of the QOF would be helpful (details are in tables but not easy to extract across multiple studies).</p>	<p>have added clarifying statements in both the executive summary and the main report. 6. Thank you. We have reorganized the structure of our KQ2 results and have added an evidence table that better clarifies the number of studies relevant to different implementation characteristics, as well as the differences between evidence and themes that arose in our KI interviews. 7. Thank you. Unfortunately there is little evidence that speaks to specific designs or number of measures that would optimize benefit and mitigate harm. Both the study evidence (<i>eg</i>, Werner & Dudley, 2009) as well as insights from our key informants suggest that factors such as patient population, organizational structure and culture, level of current performance, and organizational goals should be considered in making these decisions. Similarly, using a bottom up approach to program planning may help to identify the type of payment structure and type/number of measures that are optimal for a specific organization/health system. 8. We have added more detail about the QOF and the types of evidence related to the QOF to the introduction of the main report.</p>
--	--	---

<p>This is an excellent and comprehensive report. My suggestions are relatively minor.</p> <p>1. The report might benefit from inclusion of the results of The MA Blue Cross Blue Shield Alternative Quality Contract (See multiple articles by Zirui Song.) This structure uses a global budget with a P4P incentive program embedded, and demonstrated improvements in quality and reductions in spending. This framework seems relatively important, since many payers are moving toward a global budget as a way of holding spending in check (eg, Oregon's Medicaid transformation has a global budget with P4P embedded).</p> <p>2. Related to #1, the report might benefit from some additional discussion of how P4P is tied to the overall payment mechanism. The assumption seems to be that P4P is strictly a bonus payment generally paid on top of a FFS or salaried contract. There is not much discussion about the potential for holding providers at risk (like an ACO). I recognize that this is scope creep but a few sentences may be helpful context.</p> <p>3. Some discussion of how P4P seems to affect ambulatory primary care vs ambulatory specialist care might be helpful.</p> <p>4. One of the summary comments (page 9) seems slight at odds with earlier text on page 2.</p> <p>Page 9: "In general, P4P programs appear to have the potential to improve process of care outcomes over the short term, especially in ambulatory settings."</p> <p>Page 2 "Overall, there is low to moderate evidence that P4P programs in ambulatory settings can improve the proportion of patients receiving the care process targeted by an intervention, though these effects are typically modest, not sustained over the long term, and were inconsistent across studies.... In hospital settings, studies evaluating the Premier Hospital Quality Incentive Demonstration (HQID) and the Hospital Value-Based Purchasing (HVBP) programs in the United States report a limited effect on both processes of care and patient outcomes. However, a study evaluating the effect of P4P in the VHA on processes of care</p>	<ol style="list-style-type: none"> 1. Thank you. One article by Song appeared in the results of our search; however, we excluded the study for 2 reasons, a) it was included in the RAND report, and b) our inclusion criteria limited us to programs that were primarily described as P4P and excluded ACOs. 2. Thank you, we have added this to the revised report. 3. This is an excellent point, we have added as statement to the revision indicating that the bulk of ambulatory studies relate to primary care. 4. I believe that language may suggest that the statements are conflicting. To clarify – the statement on p.9 describes our findings – that P4P programs have the potential to improve processes of care over the short term; whereas, the statement on p. 2 describes the body of evidence as low to moderate. 5. Thank you, yes – in our revision we have reorganized our findings related to implementation to better highlight factors related to behavior and behavioral economics. 6. Thank you. We agree that this is an interesting question and worthy of study. Two included studies relate to costs and payment models; however, neither directly address the question you pose. Morgan and Beerstecher (2006) compared contract and employment status under the QOF and found that greater efficiency and higher quality were associated with GPs who were contractors. Walker et al (2010) examined the cost effectiveness of 9 QOF indicators, and found that although most indicators required only a fraction of a 1% change to be cost-effective, for some indicators improvements in performance of around 20% were needed.
---	--



	<p>found significant and sustained improvement on 6 of the 7 measures examined. Internationally, studies evaluating hospital P4P programs report generally positive effects, with a slowing of improvements or a plateau over time."</p> <p>5. It might be helpful to close with a bit more about the potential for incorporating some of the frameworks/nudges from behavioral economics into P4P programs. See <i>eg</i></p> <p>P4P4P: an agenda for research on pay-for-performance for patients KG Volpp, MV Pauly, G Loewenstein, D Bangsberg - Health Affairs, 2009</p> <p>Using the lessons of behavioral economics to design more effective pay-for-performance programs I Siva - The American journal of managed care, 2010</p> <p>6. There is very little discussion about the extent to which P4P is cost-increasing vs cost-reducing. This might be worth considering, given the interest in payment models that can reduce spending.</p>	
--	---	--

	<p>Please see attachment. I found the literature review format and summaries quite useful. While I marked the overall report as good, the literature review was excellent and is a valuable summary.</p> <ol style="list-style-type: none"> 1. Thank you for the opportunity to comment on the study. I believe the paper is well written and the following comments are to discuss some of the areas that are touched on in the body of the paper but are not as clear in the Executive Summary/Conclusions. The other contention in my comments is that the emphasis on the short term positive effects that were more apparent in ambulatory settings is somewhat at variance from my own experience working with plans, hospitals, and physicians who were individual physicians in IPAs, or otherwise better organized such as Kaiser or Sharp Medical group, and that CMS and NCQA programs such as DSRIP, Meaningful Use, etc are a form of P4P. 2. In general, as noted in the paper, what sounds like a straight forward proposition is a difficult topic, with many confounding elements. In general, the Executive Summary does a nice job of noting these, including race, ethnicity, socioeconomic status and social determinants. However, it does not place enough emphasis on the need to be aware of these elements as variables, nor does it clearly differentiate what are “ambulatory settings” – whether these are organized entities (eg. medical groups such as Kaiser Permanente, individual offices, or a mix. 3. I was pleased that the review of the literature and the summary of this in tabular form was inclusive of other countries’ experiences. This was especially useful when trying to understand responses for populations that are relatively small in the United States. As an example, the Taipei study helps with understanding issues that relate to Asian provider and patient response. 4. On page 6, lines 16-21, I would take exception to the statement that “Programs in ambulatory care have been more successful than hospital-based programs”. In my experience working with quality assurance programs for plans, hospitals, and medical groups including IPAs, I believe that the least productive programs including P4P were in the ambulatory setting in private physician settings due to lack of structure, 	<ol style="list-style-type: none"> 1. Thank you very much. In our revision, we have reorganized much of our results to better clarify differences between findings from the body of evidence vs themes that arose in our KI interviews, and to align our findings according to our framework. We very much hope that these changes will provide a much clearer presentation. With regard to findings related to ambulatory care and the organizations/programs you mentioned. The literature included in our report included only studies that were published and/or not included in the RAND report, and it is possible that some of the research related to the mentioned programs were excluded due to search date limitations. In addition we limited our scope to programs that were direct P4P programs, and did not include those that were ACOs or bundled payments. In response to your review comment, our research librarian conducted a search of the mentioned organization/program websites for unpublished studies meeting our inclusion criteria. None were located. 2. Thank you. Our designation of ambulatory vs hospital settings were based on the target of the P4P program. We have included a statement clarifying this in our revision. 3. Thank you. While we did not include studies conducted in all countries, we did include those conducted in countries in which the healthcare systems are large and the contextual settings are similar enough to generalize to the broader US and to VA settings (eg, we excluded studies conducted less developed countries such as Kenya). 4. We have removed this line from the revised report. With regard to your point. A large percentage of our ambulatory studies focused on the QOF, a program that has demonstrated success, particularly over the short term. We completely agree that ambulatory programs in the United States are incredibly heterogeneous. However, with regard to hospital based-P4P programs, the studies included in our report concluded few significant changes in process of care and patient outcome measures associated with CMS’s HQID and HVBP programs. 5. Thank you. This is an incredibly complex topic, and findings from studies are unclear with regard to the exact role that incentives play independent of other contextual and programmatic factors on improvements in quality. Given the heterogeneity in P4P programs, as well as programs that track quality metrics without financial incentives per se, it may be
--	--	---



	<p>resources, and support in the individual office based ambulatory setting. The statement should clarify if this is also included as an ambulatory setting. The effectiveness of changes of hospital behaviors are clear from the recent CMS penalties for DSRIP and such measures as hospital re-admissions. These measures have also been effective in gaining the attention of financial officers of hospitals so that there has been increased financing for improved quality programs that carry a financial consequence. These would probably be too recent to be in the literature review would seem important examples, as are such measures from CMS as the take back of Medicare payment for not complying with electronic prescribing and the meaningful use program for electronic record adoption for individual practitioners along with large groups. I didn't see this referenced in the literature review but I could have overlooked this. Adoption of EHR (meaningful use) is also a form of P4P which is not commented on. How and whether these office interventions improve or detract from quality care (especially with the poor experience with some EHRs) should call for further study.</p> <p>5. Although quality improvement cannot be shown conclusively to be based on P 4P, (Page 6, lines 26-27) I submit a P4P program helps draw attention to important quality measures. If there is no literature to support this thesis, at least this should be discussed in the summary as an important area to study. I would contend such measures at the hospital level have improved hospital care. Pointing towards guidelines for quality as versus the absence of such programs may be the most important rationale for a P4P program. As noted elsewhere (Page 66, lines 31-38) providers believe they are doing their best for patients. The criteria of a P4P program could be roadmap for such behavior, and is apart from any financial motivation.</p> <p>6. Page 7, l 36-37 appropriately discusses that P 4P programs may be of disadvantage to minority ethnic and socioeconomic groups and those who practice within those settings. However, there did not seem to be enough discussion of this important point – especially as the government programs begin to utilize P4P for reimbursement purposes. Disadvantaged groups may start at a lower baseline, and the responses to patient surveys for minority groups can vary. This also related to language and</p>	<p>safe to say that the implementation of measures serve as a roadmap for improvements in quality; however, the incentives to achieve these measures may be financial (eg, P4P) in nature, but may be linked to non-financial motivators such as public reporting.</p> <p>6. Thank you. We have added a future research need related to these topics.</p> <p>7. Thank you. Our conclusions are based on the studies we identified in both the published literature and a search of unpublished sources. As mentioned above, very few of the studies reporting outcomes related to hospital P4P programs included significant findings. While there may be significant positive effects related to P)4P in hospital settings, our conclusions were limited to studies that met our inclusion and exclusion criteria.</p> <p>8. Thank you. Yes, we do agree that provider characteristics such as the underlying payment mechanisms and other factors related to resources may play an important factor the attainment of quality. A number of studies presented in Table 9 address this issue. In studies examining the QOF, a clear trend emerged, in which larger practices showed greater improvement in the short term, and that being a contractor rather than employed was related to higher quality and greater efficiency. However, findings from studies in US and other countries were less clear, likely due to heterogeneity in programs.</p> <p>9. Thank you. We have revised the structure of the presentation of our results for KQ2, including our KI interviews. We hope that the revision better highlights some of the important themes identified through out KI interviews.</p> <p>10. Thank you. We have added a statement in our revision.</p>
--	--	--



	<p>cultural sensitivity. This is an area studied by both NCQA and CMS (with Rand).</p> <p>7. Page 9. Lines 46-47 As a conclusion, I believe lines 46-47 that “P4P programs appear to have the potential to improve process of care outlines over the short term, especially in ambulatory settings” ignores the hospital and plan experience. See my comments above. I believe there should be recognition of the CMS, NCQA, and other such programs are a form of P4P.</p> <p>8. Page 66: (Lines 31-38). The demographics did not differentiate for the US the whether providers were in fee for service, group practice, etc. Might these not provide differences in results, especially as to whether an individual physician has resources to meet the performance measures. I agree with lines 31-35 that Providers believe they are doing their best for patients.</p> <p>9. Page 70 I think lines 24-27 from KI merit appropriate emphasis in the conclusion. The entire paragraph is very important for understanding how P4P affects providers.</p> <p>10. Page 107,-108, Conclusion. This is also repeated in the Executive Summary. There should be mention about the issue of health disparities and how this may affect P4P. This was articulated well in page 7, 136-137.</p>	
--	---	--

	<ol style="list-style-type: none"> 1. Include definitions of "exception reporting", "latent variable", "LARC", "single handed practice" (several locations) 2. Table 1 needs examples for the "Outcomes" and "Need Satisfaction" rows 3. p 81: Isn't "sociodemographics" the new term-of-art? 4. p 100: make clear the concern was with use of PHQ-9 as a process measure (as opposed to an outcome) 5. p 105 L19 "gaming is inevitable" is a bit too strong in light of the evidence you present. Perhaps better to say "there is always a potential for gaming" 	<ol style="list-style-type: none"> 1. Thank you. We have defined exception reporting, long-acting reversible contraception (LARC), single handed practice throughout the report (primarily the tables). While we left reference to latent variables in Table 9, as a definition would be cumbersome within a table, we removed the use of latent variable from the report and replaced it with statistically stringent. Latent variable are construct variables that are not easily measured directly; however, are comprised of manifest (measurable) variables that can be measured. Quality of life, for example, is a latent variable – and measures don't include questions specific to quality of life, they may include those related to physical, psychological, and social function. 2. The Outcomes and Needs Satisfaction row was removed upon recommendation by L. Damschroder, who was central to the development of our model. 3. KQ3 is separated into different categories related to disparities (race/ethnicity, socioeconomic status (SES), other). We used the term SES rather than sociodemographic, as sociodemographic includes demographic factors such as race/ethnicity; whereas, SES refers more to social class, and includes a combination of income, education, and occupation. The indices included within this subcategory of KQ3 include these factors. 4. Thank you – we have replaced "gaming is inevitable" with "there is always a potential for gaming" in both the executive summary and the main report.
--	--	--

	<p>This report is a big step forward and has valuable information. The framework is unique in considering all levels of the system - program, health system/context, and embedded individuals. I have embedded comments/suggestions within the pdf file. Within context of those comments/edits I have 2 overarching comments:</p> <ol style="list-style-type: none"> 1. The framework is presented in methods but then is apparently never used again <i>eg</i>, to abstract info from studies or to organize findings 2. The summary reads like a giant laundry list with little connection between sections - though they rely heavily on one another (<i>eg</i>, the first section on programs must be interpreted within context of information presented in all of the remaining sections). Policy implications are not as coherent and actionable as they could be. Summary tables and borrowing structure from the framework would help considerably. <p>Bottom line: the report is "good" and has the potential to be "excellent"</p>	<p>Thank you for the thorough comments/suggestions. We have taken a close look at all of them, and have implemented many of your suggestions. With regard to your overarching comments below:</p> <ol style="list-style-type: none"> 1. We have revised the report to better integrate the framework and have organized KQ2 specifically around the framework. 2. Thank you. Yes, included in our revision is a summary table organized around the framework.
	<p>Regarding question 3 above relating to possible overlooked studies:</p> <p>There are quite probably more studies and papers on this topic but none of any import that would change the findings of this report that I am aware of. This overview of the literature appears to be a sufficiently broad net, encompassing public and private organizations, and multiple countries and cultures. It also includes other broad search studies, <i>eg</i>, Rand, that have undertaken similar exhaustive searches of the effects of P4P. I am satisfied that even if it doesn't cover the entire universe of existing studies on this topic that this is a very large and diverse subset and therefore very credible as a guide to both application of P4P incentives and future research on the subject.</p>	<p>Thank you for your feedback.</p>

	<p>Several comments:</p> <ol style="list-style-type: none"> 1. Add more information regarding the methodology used to select the KIs - p. 2; lines 10-15. 2. In terms of future research needed: more emphasis on rural/underserved populations as well as social determinants of health, health disparities, and the importance of patient self-reporting of exclusion. 3. Page 13, Figure 1. - External factors should be more explicitly defined; research should be included that describes the influence of public policy and the policy formulation process (state, local, federal) on P4P program design (incentives/disincentives), processes, public resource allocation, and health outcomes. 4. The theme of transparency is echoed in the report (p. 70), describing the UK's use of NICE to manage indicators and involve all stakeholders throughout the process, using a "bottom-up" approach. More research is needed that is focused on stakeholder involvement, including different levels of providers and their roles/training, and the impact on patient health outcomes related to P4P in the US 5. Common themes that emerged among KIs (p. 104) regarding policy implications, specifically the inclusion of public reporting in tandem with P4P, is essential. Methodologies could be proposed that would help delineate the value of each in quality improvement. 	<ol style="list-style-type: none"> 1. Thank you for your feedback. We have expanded the description of methods used in our KI interviews in both the executive summary and the main report. 2. Thank you. We have added additional information to the methods section. 3. We have revised the description of Outer Setting in the framework to include social norms, federal, state, and local policies. With regard to research in area, unfortunately we did not identify any studies targeting the influence of these factors. 4. Thank you. Yes, we absolutely agree, and have added it to the revised report. 5. Thank you. This is an important issue, and we have highlighted this topic as an important future research need.
--	--	--